

ドイツ語の多読支援環境の構築

田中 省作 田畑 義之

九州大学 情報基盤センター 研究部

1 はじめに

多読 (extensive reading) は、外国語学習において非常に有効な学習法の一つである。しかし、多読には学習者が一定の語彙知識や文法知識を備えている必要があり、例えば、大学におけるドイツ語やフランス語といった初習外国語では、直ちに実践することは難しい。そこで、我々は、ドイツ語を対象として、未知語や成句表現などに予め訳語情報を埋め込んだ訳語情報付き文書を多読の教材に使用している [4]。この文書は HTML で記述され、学生は Web ブラウザを通して文書を読み進める。一方、この訳語情報付き文書を教師が作成するには、HTML に関する知識が必要となること、文書が大規模になると訳語情報の埋め込みの作業量が大きな負担になること、といった点で問題があった。そこで、我々はネットワーク上に、この訳語情報付き文書の作成支援システムを構築した。本稿では、その概要を述べる。

2 初習外国語と多読学習

多読は、多量の文書を細かい内容にとらわれず、要点を掴み内容を楽しむ読み方である。近年、この多読が外国語学習に効果があることが報告されており、外国語の講義や演習に取り入れられることも多くなっている。この多読の対象とする文書は、語彙と文法に関しては学習者の言語能力内であることが前提となる。そうでなければ、文書を読み進めていく際に、辞書引きなどで読みが中断されることになる。そういった意味では、大学におけるドイツ語やフランス語といった初習外国語では、直ちに多読を実践することは難しい。そこで、従来は、学習者の習得レベルに応じた段階別読本 (graded readers) を多読の教材に用いることが多かった。この段階別読本は、レベルに応じて使用語彙や構文を制限し、原文を人工的に書き直したようなものである。しかし、このような教材を作成するには多大な労力を要し、実際、十分な数が提供されていない。その結果、学習者の多様な興味に十分対応できていない。そこで、本稿では、段階別読本ではなく、ドイツ語の実文書を多読の対象とすることを考える。

3 Web を活用した多読支援環境

3.1 多読用の訳語情報付き文書

我々は、多読の教材として、ドイツ語の実文書に対して、学習者が読み進めていく上で障害となるような単語や成句表現に訳語情報を埋め込んだ訳語情報付き文書を用いている [4]。この文書は、HTML で記述され、訳語情報も HTML のソースの一部に埋め込まれている。よって、学習者は Web ブラウザさえ使用できる環境にあれば読むことができる。文書がローカルに置かれていれば、ネットワークに繋がっている必要もない。Web ブラウザでこの文書を開くと、訳語情報が埋め込まれている表現は色付きになっており、その表現の上にマウスを合わせると、サブウィンドウが開いて訳語情報が表示される (図 1)。

表示される訳語情報は、単なる辞書項目の羅列ではなく、その文脈下での意味を端的に表したもので、予め教師が記述したものである。外国語の文書を読む場合、多義語といった文脈に依存して意味が決まる表現に対して、可能な意味の中から適切な意味を選び出す力を養うことも、もちろん重要である。だが、我々は、多読の場合は、むしろ読みが中断されることなく、先にどんどん読み進むことの方が重要であると考えている。

次に、訳語情報付き文書と学習者の読みに対する動機付けとの関係を考えておく。文献 [1] によると、外国語で読もうとする動機付けの主要な要因として、

1. 教材が学習者にとって興味深い (how interesting; attractiveness).
2. 教材の難しさが学習者の言語レベルに対して適切である (linguistic level).
3. 好きなときに読める利便性がある (availability).

が挙げられている。訳語情報付き文書は Web ブラウザさえあれば利用できるし (要因 3.)、訳語情報によって学習者と実文書間の言語レベルのギャップをある程度埋めることができる (要因 2.)。そして、実文書から作成するので学習者の多様な興味に対応できると考えられる (要因 1.)。

また、この訳語情報付き文書は、段階別読本とは異なる

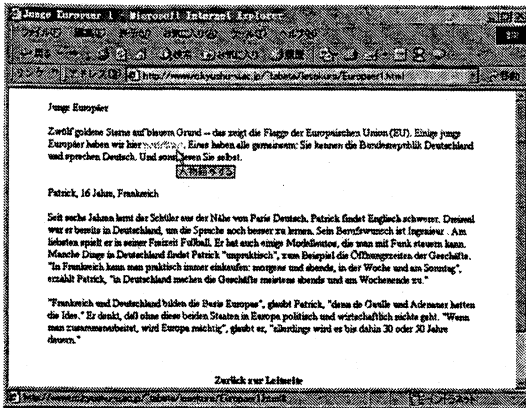


図 1: HTML の訳語情報付き文書 (“porträtiert” に対して「人物描写する」という訳語情報が表示されている)

り、実際にネイティブが読むものと同じ英文書を読破することになる。これは、学習者の読みに対する大きな動機付けになる。

一方、コンピュータ上で文書を読むということに関して、ある種の抵抗感があり、動機付けを下げる要因になるのではないかという危惧があった。だが、実際に訳語情報付き文書で多読を実践したドイツ語読解セミナーの受講者 18 名に対して行ったアンケートでは、14 名の学生が問題無いと回答した¹。

訳語情報付き文書を作成する観点からは、次のような問題があった。

- 教師に HTML に関する知識を要求すること。
- 文書が大規模になると訳語情報埋め込みの作業量が、大きな負担となること。

そこで、これらの作業を軽減する訳語情報付き文書の作成支援システムを構築した。

3.2 訳語情報付き文書の作成支援システム

本システムは、Web サーバ上に構築されており、ネットワーク経由で Web ブラウザから操作することになる。今回、構築した訳語情報付き文書の作成支援システムの流れは次のようになる。

¹ 毎年、読解セミナーの最後の講義で、アンケートを実施している。本文の記述は 1998 年に実施したものによる。本年度 (2001 年度)・後期の読解セミナーでは同様のアンケートに対して受講者 4 名全員が問題無いと回答した。



図 2: ドイツ語文書の入力

Step 1. ドイツ語文書を入力する (図 2)。

Step 2. 入力されたドイツ語文書を形態素解析し、システムがもつ独和对訳辞書 (3.3 節) を基に訳語情報を付与する。

Step 3. Step 2. の結果を教師に提示し、訳語情報のチェックを行う (図 3)。

Step 4. 訳語情報付き文書を生成する。

Step 1. ドイツ語文書の入力

作成支援システムの Web ページを開き、図 2 のフォーム部分にドイツ語文書を入力する。もちろん、普通にタイピングして入力してもよいし、ドイツ語の Web ページなどからコピー&ペーストするのも構わない。ここでのインデントは、最終的に生成される訳語情報付き文書にも反映される。

Step 2. 形態素解析と訳語情報の自動付与

Step 1. で入力されたドイツ語文書を形態素解析し、形態素毎にシステムが有す独和对訳辞書の訳語情報を付与する。なお、ドイツ語文の形態素解析には TreeTagger を

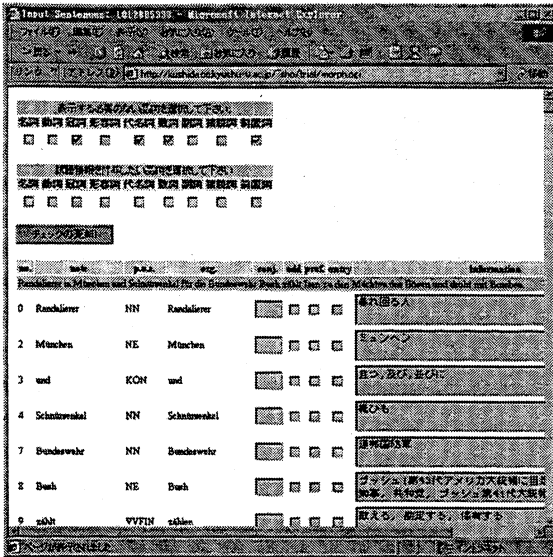


図 3: 教師による訳語情報の確認・訂正・追加

使用した²。ただし、TreeTagger は後述する分離動詞における分離前綴りの認定は行いが、基本動詞部分との対応関係までは求めない。分離前綴りと基本動詞部分の解析部分については、自前で作成した。

Step 3. 教師のチェック

Step 2. で各形態素毎に付与された訳語情報を教師に提示し、教師は、次のような操作を行う。

- 訳語情報を埋め込む形態素を選択する。
- システムが付与した訳語情報が適切でない場合や冗長な場合など、訳語情報に修正を加える。
- 修正した訳語情報を、システムの独和对訳辞書に反映させるかどうかを選択する。
- 成句表現を指定する³。

この Step 3. での訳語情報の修正は、システムの独和对訳辞書に加えることができる。更新操作を行うことで、その情報は直ちに作成中の文書の訳語情報にも反映される。

このように、教師は、HTML を全く記述することなく一貫して訳語情報に関する作業に集中できる。日本語、

²<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>

³現在のシステムでは、成句表現の自動検出は行っていない。

ドイツ語が混合することになるが、文字コードについても意識する必要はない。

Step 3. では形態素解析した結果とシステムの独和对訳辞書に従って付与された訳語情報を教師がチェックすることになる。これは教師にとって大変な作業に思われる。しかし、独和对訳辞書が十分整備されれば、実作業としては訳語情報から不適切、冗長な情報を削除するだけの作業である。これはさほど大きな負担ではない。現状の独和对訳辞書では、大体 5~6 割の形態素に対しては訳語情報が付与される。

3.3 本システムの独和对訳辞書について

本システムで作成される訳語情報付き文書は、元の文書の著作権が許せばそのまま本システムにストアし、共有、公開したいと考えている。そのため、本システムの独和对訳辞書は、市販のものではなく独自に整備したものを使用している。

内容としては、ドイツ語表現に対して対訳と適宜補助情報を加えたもので、多読では不要となる発音情報や語源情報といったものは無い。詳細は割愛するが、これらの辞書は一から作成するのはなく、既存の電子化資源、特に英語をピボットとしてフリーの独英辞書と英和辞書からの合成を試みている [5]。もちろん完全に自動合成することは不可能で、非常に信頼度が高く合成できる部分については機械的に、それが困難な部分に人手をかける、というスタンスで整備を進めている。

さらに、本システムから得られる訳語情報に関するチェック情報を用いて、独和对訳辞書は動的に精練される。特に通常の辞書では弱い固有名詞や新語、時事用語に強い辞書を構築できると思われる。例えば、「Gehard Schroder」に対して「ゲアハルト・シュレーダー (現ドイツ首相)」といった通常の辞書には記載されないような説明的な情報も時事に合わせて記載される。

現在、多読用の独和对訳辞書の登録語数は、37,625 語である。

4 関連アプリケーションとの比較

現在、日本語をベースとしたドイツ語学習に関するアプリケーションは、あまり充実しているとはいえない。ドイツ語文の多読に特化したものはなく、読解支援のものも主に単語の辞書引きの自動化に重点が置かれている。ここでは、ドイツ語を対象とした読解支援のアプリケー

ションとしてテクノクラフト社⁴の Roboword を例に挙げ、簡単に本システムとの比較、検討を行う。

Roboword は、マウスが指す単語の辞書項目を補助的に表示するものがある。しかし、表示される情報は、辞書項目の羅列であり、読みには不要の発音情報、語源情報といったものまで表示する。しかも Roboword の辞書は、通常の紙の辞書を単純に電子化したもので、それによる弊害もある。例えば、“Minderheiten” に対しては「Minderheit の複数形」という情報が表示されるだけである。本システムは、教師のチェックを一度受けることによって、訳語情報が、その文脈下で適切な訳語であることを保証することになる。

また、Roboword はドイツ語文を形態素解析をせずにパターン・マッチングで単語の品詞認定を行うため⁵、分離動詞などには全く対応できない。ドイツ語の分離動詞というのは、前置詞（分離前綴り）と動詞（基本動詞部分）から成る動詞で、文中では分離前綴りと基本動詞部分が分離して出現する。例えば、“an|kommen”（到着する）という分離動詞⁶は、分離前綴り“an”と基本動詞部分“kommen”が、次のように現われる。

Sie kommt morgen in Berlin an.
(彼女は明日ベルリンに到着する。)

単純にパターンだけを参照したのでは、“kommen”がそれだけで動詞なのか、分離動詞の基本動詞部分なのか、“an”が前置詞なのか、分離前綴りなのか、ということが決定できない。その結果、Roboword では、“kommt”に対しては、動詞“kommen”の辞書項目が、“an”に対しては前置詞“an”の辞書項目が表示されることになり、誤った情報を提示することとなる。一方、本システムでは、一度形態素解析を施し、分離前綴りと基本動詞部分の対応関係まで求めるため、システムが与える訳語情報は、基本動詞部分の“kommt”に対しては「an|kommen: 到着する」、分離前綴りの“an”に対しては「ankommen の分離前綴り」となる⁷。同様に、“weiß”が、形容詞“weiß”（白い）なのか、動詞“wissen”（知っている）の活用したものなのか、といった場合も Roboword では十分に対応できない。

⁴<http://www.technocraft.co.jp/>

⁵ドイツ語は、英語同様、分かち書きされておき形態素の認定自体は容易、しかも名詞は大文字で書き出すといった規則のため、品詞認定はパターン・マッチングだけでもある程度の精度が得られる。

⁶分離動詞を記述する場合、分離前綴りと基本動詞部分が分かるように“|”を挿入する。

⁷現在は、このような形態で訳語情報を与えているが、別手段としては、分離前綴り（基本動詞部分）にマウスが当たると対応する基本動詞部分（分離前綴り）がハイライトされるように HTML を直すことも考えている。

5 おわりに

本稿では、多読のための HTML の訳語情報付き文書と、訳語情報付き文書の作成支援システムについて述べた。ここではドイツ語を対象としたが、基本的にはドイツ語固有の処理を置き換えれば、他言語でも同様のアプローチが可能である。

今後は、本システムを公開し、実際に全国のドイツ語教師に使用して頂き、システムそのものに対する評価を行い、修正を加えていく予定である。また、形態素解析に加え統語解析の適用を検討しており、構造に関する情報や成句表現への訳語情報埋め込みの効率化が期待される。

作成された訳語情報付き文書については、著作権に問題がないもの⁸については共有し、多読学習に利用していく。本来、多読の際の教材は学習者が各自の興味に従って選択できるのが望ましく、共有される訳語情報付き文書が増えれば、学習者自らが自由な時間に読み進められるような形にしたいと考えている。

謝辞

ドイツ語文の形態素解析に TreeTagger を使用させて頂きました。Stuttgart 大学の Helmut Schmid 博士に感謝致します。

参考文献

- [1] Day, R. R. and Bamford, J.: Extensive reading in the second language classroom, Cambridge University Press (1998).
- [2] 門田修平, 野呂忠司: 英語リーディングの認知メカニズム, くろしお出版 (2001).
- [3] Schiller, A., Teufel, S. and Stöckert, C: Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS, Seminar für Sprachwissenschaft (1995).
- [4] 田中俊明, 田畑義之: マルチメディア時代のドイツ語教育, 九州大学出版会 (2000).
- [5] 田中省作: 英語を介した独和对訳辞書合成における対訳候補の絞り込み, 第 54 回電気関係学会九州支部連合大会, pp. 627 (2001).

⁸例えば、Project Gutenberg で収集された文書など。
<http://promo.net/pg/>